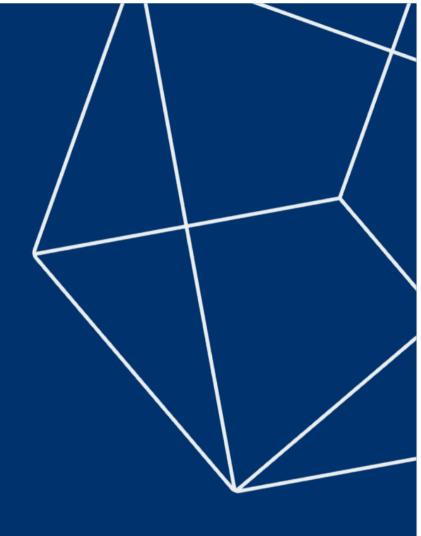# Introductions
& where I'm at

Gabriel Mateus Bernardo Harrington
Research Associate
Rebecca Sims group
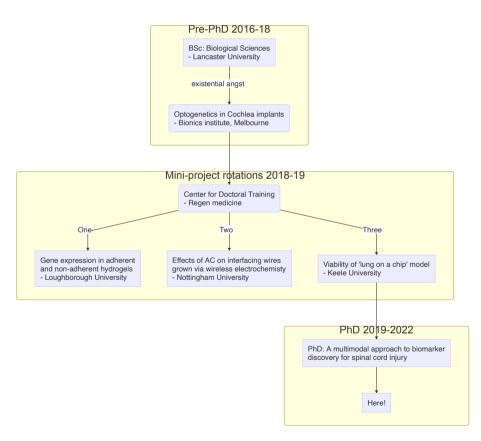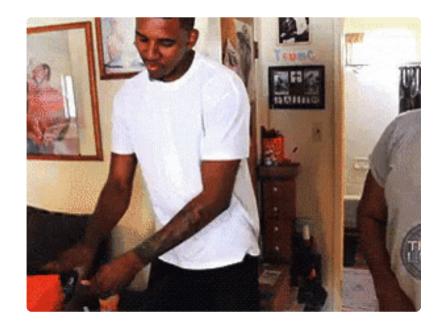**2022-05-03 (updated: 2022-04-29)**

Background

Cohort database

- Series of cohorts Cardiff has been involved in recruiting
- Rich phenotypic data
  - Problem: data not in a structure very amenable to analysis
  - Solution: create relational database to house data in a structure and well controllable manner
  - Used the free and open-source PostgreSQL
- General recommendations:
  - Spreadsheets are rubbish
  - Consider using relational databases for anything you actually care about, especially longer term projects and/or ones with lots of people inputting data

Whole exsome sequence data

- Inherited from Didi, she had done a lot of the initial file processing
- Currently grappling with VCF file annotation via Hail
- Sample IDs also not matching phenotype data very well, see previous point regarding spreadsheets being rubbish
  - Building a lookup table, pending the remaining data coming from the field team
- Data soon™ (I hope)

Reproducible slides

- I made a little R package with a template for making slideshows with the DRI theme via RMarkdown/xaringan
- Nice ability to easily host slides online, see some of my presentations here
- Package can be found in this GitHub repo: H-Mateus/dri.xaringan.theme
- An example presentation using the DRI (in addition to what you're looking at) can be found here

Thanks for listening

Gabriel Mateus Bernardo Harrington
Research Assocaite

These slides can be found from the QR code below, or at this address:
https://h-mateus.github.io/presentations/gregynog_2022-04-28/index.html#1